programa final

Encontro Nacional de Produtores e Usuários de Informações Sociais, Econômicas e Territoriais

INFORMAÇÃO PARA UMA SOCIEDADE MAIS JUSTA

III Conferência Nacional de Geografia e Cartografia

IV Conferência Nacional de Estatística

Reunião de Instituições Produtoras
Fórum de Usuários
Seminário "Desafios para Repensar o Trabalho"
Simpósio de Inovações
Jornada de Cursos
Mostra de Tecnologias de Informação

27 a 31 de maio de 1996
Rio de Janeiro, RJ BRASIL

IBGE 60 anos 1936-1996

Uma das maneiras de olhar o ofício de produzir informações sociais, econômicas e territoriais é como arte de descrever o mundo. Estatísticas e mapas transportam os fenômenos da realidade para escalas apropriadas à perspectiva de nossa visão humana e nos permitem pensar e agir à distância, construindo avenidas de mão dupla que juntam o mundo e suas imagens. Maior o poder de síntese dessas representações, combinando, com precisão, elementos dispersos e heterogêneos do cotidiano, maior o nosso conhecimento e a nossa capacidade de compreender e transformar a realidade.

Visto como arte, o ofício de produzir essas informações reflete a cultura de um País e de sua época, como essa cultura vê o mundo e o torna visível, redefinindo o que vê e o que há para se ver.

No cenário de contínua inovação tecnológica e mudança de culturas da sociedade contemporânea, as novas tecnologias de informação - reunindo computadores, telecomunicações e redes de informação - aceleram aquele movimento de mobilização do mundo real. Aumenta a velocidade da acumulação de informação e são ampliados seus requisitos de atualização, formato - mais flexível, personalizado e interativo - e, principalmente, de acessibilidade. A plataforma digital vem se consolidando como o meio mais simples, barato e poderoso para tratar a informação, tornando possíveis novos produtos e serviços e conquistando novos usuários.

Acreditamos ser o ambiente de conversa e controvérsia e de troca entre as diferentes disciplinas, nas mesas redondas e sessões temáticas das Conferências Nacionais de Geografia, Cartografia e Estatística e do Simpósio de Inovações, aquele que melhor enseja o aprimoramento do consenso sobre os fenômenos a serem mensurados para retratar a sociedade, a economia e o território nacional e sobre as prioridades e formatos das informações necessárias para o fortalecimento da cidadania, a definição de políticas públicas e a gestão político - administrativa do País, e para criar uma sociedade mais justa.

Simon Schwartzman
Coordenador Geral do ENCONTRO

Promoção

Fundação Instituto Brasileiro de Geografia e Estatística
IBGE
Fundação Instituto Brasileiro de Geografia e Estatistica
IBGE
Associação Brasileira de Estudos Populacionais
ABEP

Co-Promoção

Associação Brasileira de Estatística
ABE
Associação Brasileira de Estudos do Trabalho
ABET
Associação Brasileira de Pós-graduação em Saúde Coletiva
ABRASCO
Associação Nacional de Centros de Pós-graduação em Economia
ANPEC
Associação Nacional de Pós-graduação e Pesquisa em Ciências Sociais
ANPOCS
Associação Nacional de Pós-graduação e Pesquisa em Geografia
ANPEGE
Associação Nacional de Pós-graduação e Pesquisa em Planejamento Urbano e Regional
ANPUR
Sociedade Brasileira de Cartografia
SBC

Apoio

Federação das Indústrias do Estado do Rio de Janeiro
FIRJAN
Academia Brasileira de Letras
ABL
Conselho Nacional de Pesquisas
CNPq
Financiadora de Estudos e Projetos
FINEP
Revista Ciência Hoje

Institutos Regionais Associados

Companhia do Desenvolvimento do Planalto Central
CODEPLAN (DF)
Empresa Metropolitana de Planejamento da Grande São Paulo S/A
EMPLASA (SP)
Empresa Municipal de Informática e Planejamento S/A
IPLANRIO (RJ)
Fundação Centro de Informações e Dados do Rio de Janeiro
CIDE (RJ)
Fundação de Economia e Estatística
FEE (RS)
Fundação de Planejamento Metropolitano e Regional
METROPLAN (RS)
Fundação Instituto de Planejamento do Ceará
IPLANCE (CE)
Fundação João Pinheiro
FJP (MG)
Fundação Joaquim Nabuco
FUNDAJ (PE)
Fundação Sistema Estadual de Análise de Dados
SEADE (SP)
Instituto Ambiental do Paraná
IAP (PR)
Instituto de Geociências Aplicadas
IGA (MG)
Instituto de Pesquisas Econômicas, Administrativas e Contábeis
IPEAD (MG)
Instituto do Desenvolvimento Econômico Social do Pará
IDESP (PA)
Instituto Geográfico e Cartográfico
IGC (SP)
Instituto de Apoio à Pesquisa e ao Desenvolvimento "Jones dos Santos Neves"
IJSN (ES)
Instituto Paranaense de Desenvolvimento Econômico e Social
IPARDES (PR)
Processamento de Dados do Município de Belo Horizonte S/A
PRODABEL (MG)
Superintendência de Estudos Econômicos e Sociais da Bahia
SEI (BA)

Coordenação Geral

Simon Schwartzman

Comissões de Programa

Confege

César Ajara (IBGE)
Denizar Blitzkow (USP)
Jorge Marques (UFRJ)
Lia Osório Machado (UFRJ)
Mauro Pereira de Mello (IBGE)
Speridião Faissol (UERJ)
Trento Natali Filho (IBGE)

Confest

José A. M. de Carvalho (UFMG)
José Márcio Camargo (PUC)
Lenildo Fernandes Silva (IBGE)
Teresa Cristina N. Araújo (IBGE)
Vilmar Faria (CEBRAP)
Wilton Bussab (FGV)

Comissão Organizadora

Secretaria Executiva - Luisa Maria La Croix
Secretaria Geral - Luciana Kanham
Confege, Confest e Simpósio de Inovações
Anna Lucia Barreto de Freitas, Evangelina X.G. de Oliveira, Jaime Franklin Vidal Araújo, Lilibeth Cardozo R.Ferreira e Maria Letícia Duarte Warner
Jornada de Cursos - Carmen Feijó
Finanças - Marise Maria Ferreira
Comunicação Social - Micheline Christophe e Carlos Vieira
Programação Visual - Aldo Victorio Filho e Luiz Gonzaga C. dos Santos
Infra-Estrutura - Maria Helena Neves Pereira de Souza
Atendimento aos Participantes - Cristina Lins
Apoio
Andrea de Carvalho F. Rodrigues, Carlos Alberto dos Santos, Delfim Teixeira, Evilmerodac D. da Silva, Gilberto Scheid, Héctor O. Pravaz, Ivan P. Jordão Junior, José Augusto dos Santos, Julio da Silva, Katia V. Cavalcanti, Lecy Delfim, Maria Helena de M. Castro, Regina T. Fonseca, Rita de Cassia Ataualpa Silva e Taisa Sawczuk

Registramos ainda a colaboração de técnicos das diferentes áreas do IBGE, com seu trabalho, críticas e sugestões para a consolidação do projeto do ENCONTRO.

GEOGRAPHY AND SOCIAL, ECONOMIC, AND TERRITORIAL INFORMATION:

APPLICATIONS AND BENEFITS OF TIGER AND THE DECENNIAL CENSUS TO DATA ANALYSIS

A Paper Presented at
The Brazilian National Symposium of Producers and Users of Social, Economic, and Territorial Information

Rio de Janeiro, Brazil
May 27 - 31, 1996

Robert W. Marx
Associate Director for Decennial Census
Bureau of the Census
United States Department of Commerce
Washington, DC 20233-0170

Voice: 301-457-2131
FAX: 301-457-1902
Internet: rmarx@census.gov

ABSTRACT

The TIGER System means different things to different people. During the past four years, the TIGER System has fulfilled the precensus geographic support functions for which the Geography Division of the United States Census Bureau designed it. The TIGER System also has provided the geographic and cartographic products needed to complete the tabulation of the data collected during the 1990 decennial census and make those data useful to the numerous constituencies that carry out the myriad tasks that define our lives. Simultaneously, work is underway to define a framework for the future of this bold new product -- and this future looks bright. Similar planning is going on in offices and institutions across the United States and around the world. This is true especially in the context of geographic information system (GIS) applications involving the digital geographic and cartographic products of the TIGER System and the demographic data products of the 1990 decennial census.

INTRODUCTION

The TIGER System means different things to different people. The focus of most earlier articles about the TIGER System has been on its development and use within the United States Bureau of the Census (Marx et al 1990); these articles have described a series of massive geographic and cartographic production operations performed in conjunction with the 1990 decennial census of the United States. This article will summarize the salient points of that history and discuss several aspects of where the future might lie for the TIGER System, both within the Census Bureau and as a tool for geographers, cartographers, and others developing automated geographic and cartographic systems. It also will describe the potential for using this new system -- and the associated 1990 census statistical data -- to perform spatial analysis using geographic information system (GIS) technology.

WHAT IS THE TIGER SYSTEM?

The Geography Division, one component of the Census Bureau, developed the TIGER System in response to a major goal the agency set in 1981:

> "to automate the full range of geographic and cartographic support processes in time to serve the data collection, tabulation and dissemination needs of the 1990 decennial census -- the Bicentennial Census of the United States."

KEY WORDS: automated cartography, geographic information systems, TIGER System, U. S. Census Bureau, decennial census, future

This decision gave the Geography Division six short years to build a computer data base containing every known street and road in the United States (a task accomplished only with significant assistance from the United States Geological Survey -- USGS), the name of each, and the range of address numbers located along each segment of every street in the 345 largest urban cores of the United States; six short years to include all the railroads in the United States along with all significant water features and their associated names; and six short years to enter and verify the boundaries, names, and numeric codes for all the geographic entities used by the Census Bureau to tabulate the results of both the 1980 and 1990 decennial censuses.

This development task is complete; the overall goal, met -- and on time! Is the TIGER data base perfect? Of course not! But most of the information it contains is correct and it is much more up-to-date than the information on the traditional maps and in the GBF/DIME-Files the Census Bureau used for the 1980 census. More importantly, now all this information is in the computer! As a result, the Census Bureau can enter changes easily as people identify needed updates.

HOW HAS THE CENSUS BUREAU USED THE TIGER SYSTEM?

At this juncture, the Census Bureau has relied on the TIGER System to compile, validate, and geographically assign (geocode) its housing unit and group quarters address list for the 1990 decennial census; used it to provide the geographic and cartographic products needed to complete and evaluate the data collection operations of that census; used it to support the numerous functions needed to complete the tabulation of the collected data and make those data useful to the many constituencies that by law, or for other purposes, rely on the results of the decennial census to carry out the myriad tasks that influence our daily lives (Tomasi 1990); and has made available a "public" version of the TIGER data base that matches, in exact geographic detail, the geographic codes associated with the statistical data products of the 1990 census -- to support the GIS applications of people and organizations that will make use of these data products as the Census Bureau released them.

WHAT LEVEL OF GEOGRAPHIC DETAIL DOES TIGER PROVIDE?

The Census Bureau uses a basic geographic hierarchy for most of its statistical programs (see Figure 1). The entities in this hierarchy generally share a "nesting" relationship; that is, each major entity generally contains multiple smaller entities.

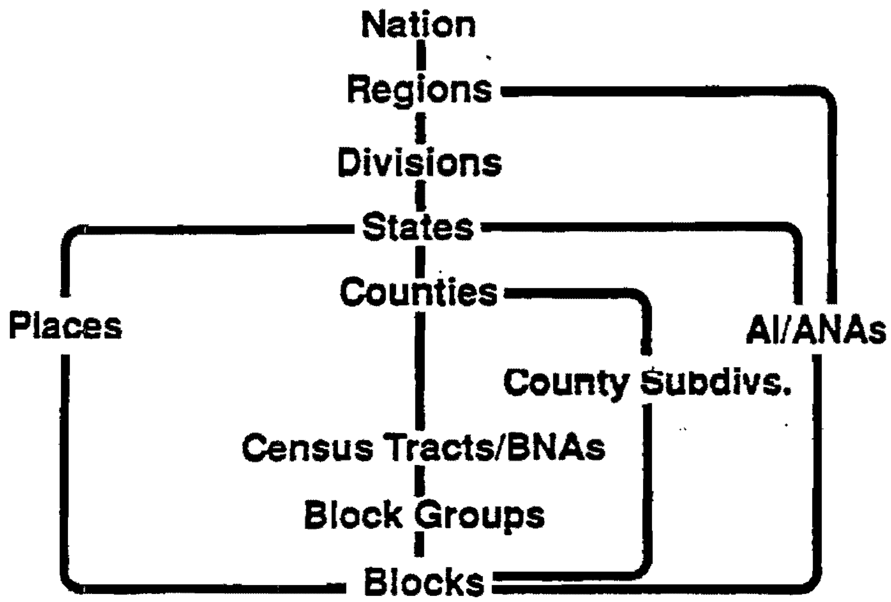

Figure 1: The Census Bureau's Basic Geographic Hierarchy

The "Higher Level" Entities

The hierarchy starts with the United States -- the entirety of the 50 states plus Washington, DC. Dividing the United States, there are 4 regions (that are groups of states) that are divided into 9 divisions (that also are groups of states). Then there are the 50 states and Washington, DC. The Census Bureau built the TIGER data base to include, in addition, Puerto Rico and the other outlying areas in which the Census Bureau helps conduct the censuses of population and housing -- American Samoa, Guam, the Northern Mariana Islands, Palau, and the Virgin Islands of the United States.

Dividing the states and those other statistically-equivalent entities, there are more than 3,200 entities that people generally call counties. Most of these entities cover fairly large geographic areas, which means that they provide fairly coarse "resolution" for the characteristics for the people, housing units, farms, and businesses of the United States -- useful for analysis mostly on a national or global scale. These counties are extremely important for data analysis purposes because the Census Bureau also makes statistical data available for them from the economic and agriculture censuses it conducts every 5 years; censuses that include topics as diverse as manufactures, retail trade, service industries, wholesale trade, construction industries, mineral industries, transportation, enterprise statistics, minority and women-owned

businesses, and all sorts of agricultural information. Many other governmental agencies and private organizations also collect and report data for counties, resulting in a "rich" data series for them.

At the next level in the Census Bureau's geographic hierarchy there are more than 60,000 units of local government and statistically equivalent entities subdividing those 3,200-plus counties -- townships, cities, villages, census designated places, and so forth. The relationship of places to the hierarchy is not quite as neat and tidy in all instances, as places do not provide "wall-to-wall" coverage the way the previously discussed entities do. In terms of data cells, these "governmental" tabulation units provide more than a twenty-fold increase in entities over counties. This still offers fairly coarse resolution in a GIS context because many subcounty governments also cover very large areas. The good news is that demographic data are available for all these entities and economic data are available for the most populous of them.

The "Lower Level" Entities

To zoom in closer, the hierarchy includes more than 62,000 census tracts -- and their cousins, the block numbering areas -- for the 1990 census. There has been a big increase in the number of these entities since the 1980 census because these entities now cover the entire United States and its territories. These census tracts and BNAs "cut-up" the large governmental units into geographic "chunks" that average about 4,000 people. This starts to make the resolution of the 1990 census tabulation entities much more useful in most GIS applications. As an example, at the county subdivision/place level in the Census Bureau's geographic hierarchy, there is one entity called "the City of Los Angeles;" at the census tract level, Los Angeles has 737 entities or which the Census Bureau tabulates detailed statistical data. These census tracts are relatively small geographic areas within the much larger City of Los Angeles.

The census tracts and BNAs are themselves subdivided into approximately 229,000 block groups that further segment the governmental units for purposes of statistical data presentation. At both the census tract/BNA and the block group levels (as at all previously discussed "higher" levels), a GIS user has access to the full range of decennial census statistical data -- those data collected from every person and about every housing unit and those data collected from only a sample of the population and their housing units.

Finally, for the 1990 census the Census Bureau identified and numbered more than 7 million census blocks -- people polygons -- covering every portion of the United States. The Census Bureau has tabulated the data it collected from every person and about every housing unit for each of these blocks. As with the census tracts and BNAs, there has been a huge increase in the number of these entities since the 1980 census when block-level data were available primarily for the urban cores of metropolitan areas. These "millions and millions" of blocks provide a very fine-grained resolution to

the demographic data sets available from the Census Bureau. They are defined by the roads, rivers, railroads, and governmental unit boundaries that are the underlying "geometry" of the TIGER data base; the same features that already comprise at least one layer in most people's GIS.

Special Purpose Entities

The Census Bureau's geographic structure includes several other categories of entities for which it tabulates many of the statistical data items it collects; these entities provide varying levels of geographic resolution and coverage.

- Metropolitan areas generally are composed of a county containing a large city or major urban center, plus adjacent counties that are linked economically to the central one. In the New England states, metropolitan areas are composed of clusters of cities and towns instead of clusters of counties. Collectively, all the metropolitan areas comprise the "metropolitan population of the United States;" everyone outside these metropolitan areas is "non-metropolitan."

- Urbanized areas are quite different. Whereas metropolitan areas are defined as collections of governmental units -- that can vary widely in population density and areal extent -- urbanized areas are defined strictly on the basis of population density. Thus, every metropolitan area contains at least one urbanized area at its core -- some contain two or more. These urbanized areas generally cover much smaller geographic areas than do metropolitan areas -- but at a much higher average population density. A number of urbanized areas also exist in counties that do not qualify for definition as metropolitan areas. Collectively, all people in these urbanized areas, together with all people in other incorporated and census designated places of 2,500 population or more, comprise the "urban population of the United States;" everyone else is part of the "rural" population.

- The 1990 census also includes data tabulations for American Indian and Alaska Native areas, Congressional Districts, and soon will include data for ZIP Codes plus a wide variety of other entities.

WHAT MAKES TIGER USEFUL FOR GIS APPLICATIONS?

Numeric codes that match those used in the Census Bureau's statistical data products are the most important contribution of the TIGER data base in the "world of GIS." Using these codes, a GIS can link statistical data directly with the network of features bounding the millions of 1990 census blocks, and thus, to the many other data sets that increasingly are available; state and local data sets as well as the data sets that flow from the decennial, economic, and agriculture censuses (see Figure 2).

In earlier days, when data analysts made maps showing the distribution of various data items, they often performed their analysis using paper maps and colored pencils! In a GIS environment -- using the TIGER data base -- the computer does the "coloring" using the numeric codes in a TIGER Extract product as a means to link with, display, and analyze the statistical data from the decennial, economic, and agriculture censuses. These codes let the data analyst examine the characteristics of the people who occupy the land -- their houses, their farms, their businesses, and their industrial activities. They also let the GIS display these data in the context of the governments responsible for managing an area and examine the inhabitants in conjunction with other geographically distributed data sets -- soil categories, hazardous waste sites, water quality, land use/land cover, sales data, and so forth.

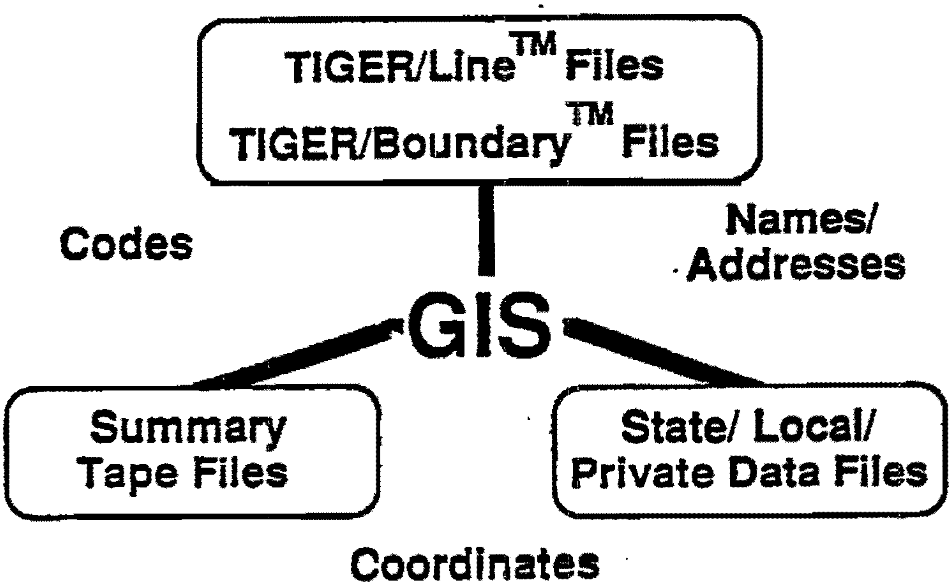

Figure 2: GIS and the TIGER System

WHAT ABOUT THE FUTURE OF THE TIGER SYSTEM?

Simultaneously with the Census Bureau's use of the TIGER System to support its 1990 census data tabulation and dissemination needs, work is underway to define a framework for the future of this bold new product; and this future looks bright. One aspect of this framework definition exercise has been to document the fact that the TIGER System and the 1990 census Address Control File are "national resources." The TIGER data base provides a computer-readable street map of the entire United States that the Census Bureau is using to support all its statistical programs. However, TIGER provides only half of what each program needs: Each census and sample

survey also needs an address list; the decennial census needs the biggest list no matter what methodology is selected for the 2000 decennial census. Further, the decennial census address list also serves as the sampling frame for the Census Bureau's demographic sample surveys.

Unlike fine wine, however, neither the TIGER data base nor the 1990 census address list improves with age. Both become more out-of-date every minute as the bulldozers keep scraping new streets out of farm fields and woodlands, and as the construction industry builds its millions of new housing units each year. In addition, these two massive files still exist separately. As things stand, Census Bureau staff must apply every needed update twice; once to the TIGER data base and once to the address list.

Another part of the framework definition exercise has been the preparation of a "geographic vision statement" to guide the decision making process when questions arise about the potential value of alternative approaches; it reads as follows:

> We will manage, in an automated format, a continually updated and increasingly accurate map base for the United States and its territories that includes the correct mailing address and geographic location of each housing unit, group quarters, farm, business, and industrial establishment. This computer file consistently will exceed the expectations of our customers, both internal and external, for current maps and correctly geocoded statistical data. It also will provide complete address lists to serve as the frame for approved statistical activities.
>
> To do this, we will seek to build partnerships with others who share our concern for having current and correct geographic and address information. Doing so will enable us to extend and improve on our earlier geographic innovations so that future products and processes will better serve the Census Bureau's respondents, customers, and partners.

A Comprehensive Address List

It appears that the nature of the Census Bureau's work will remain essentially the same in the foreseeable future. For this reason, the Census Bureau likely will continue to use address lists -- whether purchased from commercial vendors, compiled by staff in various field operations, or maintained as internal files updated through computer matches with the address lists maintained by others -- as a principle framework for its censuses and sample surveys.

To the extent that the Census Bureau obtains new or updated address lists for each of its statistical programs, the address matching function of the TIGER System will continue to play a major role. This function provides one of the most significant sources of information for updating the TIGER data base, especially in areas with structure-number/street name-style address systems. The resolution of the addresses that the TIGER System cannot

match and assign to a correct location highlights new features, feature names, and address ranges that need to be added to the TIGER data base as well as new addresses that need to be added to the Census Bureau's master address list. These additions ultimately will enhance the set of information available to do even more automated address matching and produce more effective cartographic displays.

Grid Reference Maps

A byproduct of having coordinate values that identify the location of every street intersection and end point in the TIGER data base is the capability the TIGER System offers to replace a labor intensive and error prone methodology for preparing an often requested product; a street name index referenced to a map grid. The Census Bureau used automated processes to prepare a more generalized version of this product when it created its county-based street name-to-enumerator assignment area reference listings as part of the 1990 census map production process. While using the TIGER System to prepare the maps for the printed 1990 census reports, staff inflicted a uniform grid, identified with letters and/or numbers, over the entire United States. Having done so, the software can document every grid-cell through which a street (or other mapped feature) passes, sort the resulting street name list alphanumerically, and generate the desired listing. This product will be especially useful to the sample survey operations of the Census Bureau in which the field staff must visit only a sample of housing units. The TIGER System will provide not only the map and index products that inform the field representative of the location of the sample unit, but also will provide the potential, using GIS technology, to calculate the shortest route to get him or her to the unit.

Geographic Relationship Files

Similarly, while essentially a non-cartographic function, the preparation of various geographic reference files as extracts of the TIGER data base will remain a primary function of the TIGER System. These files document the hierarchy of and relationships between and among the geographic entities used for the tabulation of Census Bureau data. They provide the geographic entity "stubs" for the printed reports and summary tape files that flow from each census and sample survey--including those that will come after the 1990 census.

Structure of the TIGER Data Base

The success of the structure used for the TIGER data base reigns as one of the most significant conceptual advances in the entire TIGER System. It provides direct linkages among the geographic entities for which the Census Bureau tabulates data. It also provides direct linkages between the object representation of those entities and the object representation of the underlying feature network that forms the cartographic base for many maps and geographic information system displays. Significantly, it does this

without the need for multiple overlays of boundary polygons typical of many automated cartographic systems.

The storage and retrieval approach used by the TIGER data base avoids all the traditional problems associated with multiple representations of a single line in many different feature (object) categories; for example, a state boundary that also is a county boundary, that also is a township boundary, that also is a city boundary, that also is a road. No matter what scale a cartographer selects for a map (graphic image), and no matter which categories of information the cartographer chooses to display, the relationships remain constant. In this way, there is never a chance that the simultaneous display of more than one category of object (for example, a boundary that follows a road) will result in slightly different alignments or different generalizations in relation to the basic underlying features.

Nevertheless, the structure of the files comprising the TIGER data base clearly warrant review based on the experience gained from the demands of 1990 census processing. Although the address matching function and the geographic reference file function of the TIGER System continue to be critical elements in planning the future applications software for the TIGER System, the records of computer hours used for each category of applications demonstrate that the most significant activity -- from a computational standpoint -- was the preparation of the cartographic products and the organization of the information in the TIGER data base to support those cartographic tasks. As a result, major attention will be given to reevaluating the original decision to avoid some redundancy in storing cartographically oriented data at the expense of computer processing costs.

Cooperation for Updates

The Census Bureau traditionally has used labor intensive map-to-map comparison techniques as the primary means to update the cartographic base features on its maps (and now in its TIGER data base). The increasing availability of automated mapping systems and GIS technology, coupled with the increasing sophistication of map makers and map users at all levels of government and in the private sector, makes the pursuit of cooperative programs and automated transfer approaches very appealing.

One of the great successes of the TIGER System to date has been the cooperative relationship that developed between the Census Bureau and the USGS. The likelihood is great that the USGS will continue to be a major cooperator and there appear to be other agencies interested as well. For example, an experiment is underway to explore cooperation with the United States Postal Service (USPS) as a means to update both the road and address information in a merged TIGER data base and nationwide address list. Another experiment in North Carolina involves cooperation among multiple agencies and organizations to develop and maintain more detailed digital cartographic files through the cooperation of a state agency, local governments, the American Association of State Highway and Transportation

Officials, the USGS, and the Census Bureau. A similar experimental study is underway in Texas involving the agencies establishing emergency service "911" telephone systems. The door to innovation remains open at the Census Bureau.

Digital Products and GIS

If the TIGER System is to be judged truly useful outside the Census Bureau, similar planning will need to be going on in offices and institutions across the United States and around the world in the countries wishing to do business in the United States. The analytical power available to study the interrelationships among diverse data sets, using the GIS technology now available in the private sector, make better understanding a very real possibility.

REFERENCES

Carbaugh, L. W., and Marx, R. W. 1990. The TIGER System: A Census Bureau innovation serving data analysts. *Government Information Quarterly*, Vol 7, no. 3, pp 285-306.

LaMacchia, R. A., Tomasi, S. G., and Piepenburg, S. K. 1987. The TIGER file: Proposed products. Paper distributed at the fall meeting of the National Conference of State Legislatures, Hartford, Connecticut.

Marx, R. W. (Guest Editor), et al. 1990. Special Content: The Census Bureau's TIGER System. *Cartography and Geographic Information Systems*, Vol 17, no. 1, pp 9-113.

Tomasi, S. G. 1990 Why the nation needs a TIGER System. *Cartography and Geographic Information Systems*, Vol. 17, no. 1, pp. 21-26.